PRECISION AGRICULTURE: TEMPORAL YIELD CORRELATION OF GRID YIELD DATA FROM HENRY COUNTY, OHIO OVER THE YEARS 1995-1998

Honors Project

Presented in Partial Fulfillment of the Undergraduate Honors Program Requirements in

the College of Food, Agricultural and Environmental Sciences of

The Ohio State University

by

Melissa R. Wright
Academic Advisor: Dr. Carl Zulauf
Project Advisors: Dr. Gary Schnitkey and Dr. Carl Zulauf

****

The Ohio State University

3 June 1999

ABSTRACT

This study calculated Pearson correlations between yield grids from three fields in Henry County, Ohio over the years 1995-1998. Field 1 was 176 acres, Field 2 was 140 acres and Field 3 was 144 acres. Overall, the study found little correlation when comparing yields for a grid with yields from previous years for that grid. However, a greater level of statistically significant correlations emerges when the information set includes the yields for previous years for the grid and surrounding grids. This finding suggests that expanding the geographical area used to explain a grid's yield may increase the predictability of a grid's yield. However, even for the later analysis, the average absolute value of the correlations is 0.14. Thus, the previous yields explain only a small amount of the variation in yields within a field. Furthermore, all but one of the significant correlations are negative, implying that the current year's grid yield is inversely correlated with yields in previous years. Thus, grid yields tend to revert to the mean grid yield for the field. In conclusion, these results call into question the economic profitability of precision agriculture; although, it must be quickly added that this is an extremely small sample over only a few years. Thus, additional work is needed that broadens the areas studied and the length of time studied.

Precision Agriculture: Temporal Yield Correlation of Grid Yield Data From Henry County, Ohio Over the Years 1995-1998

Introduction/Review of Literature

For years farmers have tried to match cropping practices to unique characteristics that occur in a field, but in general this management goal was economically unfeasible. Thus, farmers learned to treat their land with uniform, standardized techniques. However, recently introduced satellite sensing has reenergized the effort to target management practices at individual areas in fields. This effort is referred to by various names: precision agriculture, precision farming, prescription farming, site specific management (SSM), and variable rate technology (VRT). Throughout this study, the terms precision agriculture and precision farming will be used.

For the sake of this analysis, we adopt the following definition from Krill (1994): precision agriculture is "a management concept which recognizes variability within the soil and crop environment and maximizes economic production while minimizing environmental impact for a specific location." Operationally, precision agriculture begins with the division of a field into smaller units called grids. Next, soil samples are collected, the chemical composition of the soil for each unit is analyzed and maps of the soil profiles of the field are created. At harvest, yield monitors installed on the combine with a satellite dish sends and transmits signals from Global Positioning System satellites. As a result, a yield can be calculated for each grid, leading to the creation of yield maps for the field. Combining these two sets of information, the farmer can treat the field on a per grid basis by using variable rate technology to systematically vary the amount of inputs applied. Potentially, this ability to target inputs on a grid basis can decrease runoff of chemicals, reduce costs and increase yields (Weiss 1996).

Startup costs for precision agriculture technology include variable rate technology machinery, computer software, yield monitors and field analysis. These costs must be recovered over time through lower annual costs for inputs and/or higher output for farmers to adopt precision agriculture technology. A necessary condition for precision agriculture to be economically feasible is that soil and yield characteristics must exhibit sufficient yield variability. This condition has been demonstrated by on-farm experience with precision agriculture technology (Jaynes and Colvin 1997).

A second necessary condition is that yields for the same grid be correlated across years. If grid yields are correlated over time then farmers can implement management decisions directed at taking advantage of site-specific yield variation that remains stable over time. If grid yields are not correlated over time then there is no pattern to yields that can be managed by targeting production practices to specific site. In short, using uniform management practices for the entire field is optimal.

Eghball and Varvel (1998), state that "characterizing spatial and temporal variability is important in site specific studies...to evaluate the effects...on crop performance." They studied the temporal grain yield variability of seven crop sequences of corn, soybeans and sorghum over a period of twenty years to determine whether temporal or spatial variability dominated. Findings were that temporal variability dominated spatial variability. They also found that environmental conditions had a greater effect on resulting yields than did management practices.

Based on these findings, Eghball and Varvel conclude that yield maps may not be useful in making management decisions when temporal variability is great. A further potential implication is that, during adverse weather conditions (i.e. heavy rain), site specific management practices may produce highly variable results from year to year. This would cause problems in the interpretation of yield maps (Eghball and Varvel 1998).

Given this preceding discussion, the objective of this study was to evaluate the temporal correlation among yields. Three fields located in Henry County, Ohio comprised the study. Grid yield data were collected for the years of 1995-1998. Pearson correlations are calculated for yields over these four years.

The next section contains a detailed discussion on the collection and analysis of the yield data. This discussion is followed by a discussion of the Pearson correlations that can be found in tables 2-5. Lastly, conclusions and further implications are presented.

Data Collection and Description

The data for this research came from a farm located in Ridgeville Township, Henry County, Ohio. Grid yields were available from three fields on this farm for a period of four years, 1995-1998. Field 1 is 176 acres while Fields 2 and 3 are 140 and 144 acres, respectively.

The fields were mapped using a GPS system based on a Fortran program. The grids were laid out using these field maps. Each grid was one acre in size.

An "Ag Leader" yield monitor was used to measure the yields per grid. The monitor outputs many yields for each grid. The yields that fell within each grid were averaged to find the yield per grid.

The crop rotation for 1995 through 1998, respectively, was soybeans, corn, corn and soybeans for Field 1, soybeans, corn, soybeans and corn for Field 2 and corn, soybeans, corn and soybeans for Field 3 (Table 1). Also presented in Table 1 is the summary data for grid yields in the three fields for each of the four years.

The mean yields for Field 1 from 1995 through 1998 were 39.1, 135.2, 120.1 and 48.3 bushels, respectively. The highest grid yield for corn in Field 1 was 166 bushels in 1996, whereas the minimum grid yield was 90 bushels in 1997. For soybeans, the maximum grid yield occurred in 1995 at 67 bushels while the lowest in this field for soybeans was 32 bushels in 1998.

The mean yields for Field 2 from 1995 through 1998 were 39.1, 124.1, 31.6 and 139.8 bushels, respectively. For Field 2, the largest grid yield for corn was 172 bushels in 1998, and the smallest grid yield was 124 bushels in 1996. For soybeans, in Field 2, the maximum grid yield came in 1995, at 67 bushels. The minimum soybean grid yield occurred in 1997 at 30.5 bushels.

The mean yields for Field 3, from 1995 through 1998, were 124, 32.9, 130.4 and 51 bushels, respectively. In Field 3, the maximum corn grid yield was 169 bushels in 1995 while the minimum corn grid yield was 70.0 bushels. For soybeans, the largest grid yield was 112 bushels in 1998 and the smallest was 6 bushels in 1996.

The median grid yield tended to be very close to the mean grid yield. In general, the median was only approximately one bushel off of the mean. This finding implies that there was not much skewness in the distribution of grid yields.

The standard deviations for yields across grids were similar in size for both corn and soybeans. These findings imply that in a relative sense the variability of grid yields for corn is less than the variability of grid yields for soybeans, because the average grid yield for corn is 3.2 times larger than the average grid yield for soybeans. This finding is similar to Eghball and Varvel's (1998) finding that the variability of corn yields is much less compared to its mean than that of soybean yields.

Temporal Yield Analysis

To calculate temporal yield correlations, the grid yields were transformed into standardized yields using the following formula:

$$Y_{sit} = (Y_{it} - \bar{Y}_t) / G_t$$

where:

Y_{sit} = standardized yield for grid i in year t

Y_{it} = yield for grid i in year t

$\bar{Y}_t$ = mean yield for the field in year t

G_t = standard deviation of grid yields in year t

Standardizing the yields creates a variable which has a mean of zero and a standard deviation of one for each year. This transformation of the data facilitates the comparison of years that involve different crops by accounting for the difference in mean yield levels between corn and soybeans. After standardizing the yields, the CTI program in Excel97 was used to calculate the Pearson correlations and confidence intervals associated with statistical significance.

The Pearson correlations between a grid's standardized yield and that grid's standardized yield during previous years are presented in Table 2. For example, for Field 1, standardized grid yields in 1996 had a Pearson correlation with the grid's standardized yield in 1995 of -0.126. Thus, the grid yields were negatively correlated between the two years, implying that a low yield in 1995 was associated with a higher yield in 1996 while a high yield in 1995 was associated with a lower yield in 1996. The correlation was significantly different than zero at the commonly-used 90% confidence test level. This conclusion means that there is less than a one in ten chance that the correlation is due to random chance.[1]

A zero correlation implies that there is no relationship between the grid's yield in different years. In total, only three of the fifteen correlations are statistically different than zero at the 90% confidence level. They occur when comparing 1996 with 1995 in Field 1, 1998 with 1997 in Field 2, and 1997 with 1996 in Field 3. Two of these correlations are negative and one is positive. The largest correlation in an absolute value sense is 0.224, which is relatively small considering that Pearson correlations vary between -1 and 1.

Given that the correlation of grid yields between pairs of years reveals little in the way of a consistent and statistically significant relationship, a question arises as to whether additional information may improve the significance of the relationship. One potential source of additional information is the average standardized yield for previous years. For example, does averaging the standardized yield for a grid over the years 1995, 1996 and 1997 improve the correlation with the standardized yield for 1998.

Table 3 contains the Pearson correlations between the standardized yield for a grid for one year versus the average of the standardized yields of the same grid for previous years. Essentially, the same story emerges. Only one of nine correlations is statistically significant: 1997 with 1995-1996. Its correlation is negative.

Another potential source of additional information is the standardized yields for the grids surrounding a grid. This approach expands the area serving as a source of information concerning yields in later years. Thus, table 4 presents the Pearson correlations between the standardized yield for a grid for a year, and the average standardized yield for the same grid and the surrounding grids for a previous year. The number of surrounding grids are either 2, 3 or 4 depending on whether the grid was on the outside corner of the field, outside row but not at the corner of the field, or inside of the field.[2] Compared with table 2, in which only three of fifteen correlations were statistically significant at the 90% confidence level, seven of the fifteen correlations were statistically significant.

The number of significant correlation substantially exceeded the number associated with random chance. At a 90% confidence level, random chance implies that only 1.5 of the correlations should be significant (15 * 10% = 1.5). Six of the seven significant correlations were negative. This finding implies that previous yields were inversely related to yields observed during a future year.

The last source of information utilized for this study was the comparison of the grids with the average standardized yield for that grid and the surrounding grids for previous years. Table 5 presents this data. Compared with table 3, where only one out of nine correlations were statistically significant, four of the nine correlations in table 5 were statistically significant at the 90% confidence level. All of these correlations were negative.

Summary, Conclusions and Implications for Further Research

This study examined grid yield data from three fields in Henry County, Ohio for a period of four years, 1995-1998. The grids were one acre in size. The yields were standardized so that corn and soybean yields could be compared. Pearson correlations were calculated using the standardized yields.

When assessing the results of this study, it is important to keep in mind that grid yields were analyzed only for four years in three fields for one farmer in one county. Thus, the results of this study are constrained by the small sample problem. With this important caveat in mind, the following conclusions can be made based on the presented data.

When comparing yields for a grid with yields from previous years for that grid, the level of statistical significance of the correlations was limited. Only five correlations out of twenty-seven possible correlations were statistically significant at the 90% confidence level. The magnitude of the correlations was relatively small with the largest being -0.224 in table 2 for Field 3 (1997 vs. 1996).

A greater level of statistically significant correlations emerges when the information set includes the yields for previous years for the grid and the surrounding grids. The number of significant coefficients more than doubles to eleven out of twenty-seven possible correlations. In addition, the magnitude of the absolute value of the correlation coefficients is larger when the information set includes the surrounding grids. These findings suggest that expanding the geographical area used to explain a grid's yield may increase the predictability of a grid's yield.

For this expanded grid area analysis, the average absolute value of the significant coefficients is 0.23 and the average absolute value of all twenty-seven correlations is 0.14. These are relatively small correlations. To illustrate, even for the significant correlations, the previous year's yields for the grid and surrounding grids explains only 5% of the variation in the grid yields. This low level of explanatory power calls into question the economic profitability of precision agriculture when grids are laid out according to a standard geographical size. Lowenberg-DeBoer and Eghball and Varvel reach a similar conclusion.

Although the number of studies are limited, the consistency of the conclusion regarding the lack of predictable yield behavior when grids are laid out according to standardized geographical size complicates the search for a viable precision agriculture management system. It is possible that grids could be laid out according to other variables, such as soil fertility factors, topography, soil types, etc. Research by Lowenberg-DeBoer suggests that grids laid out according to these factors may be usable in a precision agriculture management system. However, this approach to precision agriculture increases both the demands for information and analysis, thus raising the startup costs associated with the adoption of precision agriculture technology.

Ten of the eleven significant correlations are negative. This means that the yields are inversely correlated. If yields are high one year, they are lower the next, and vice versa. This implies that the grid yields have a tendency to be mean reverted, i.e. they tend to revert back to the mean grid yield for the field. If additional research confirms that mean-reverting behavior is a common characteristic of a field, then management strategies that are tailored to the level of past yields for a grid must take into account this mean-reverting behavior if they are to be successful. In addition, if additional research confirms that mean-reverting behavior is a common characteristic of a field, then on-farm research must take into account this mean-reverting behavior or else inappropriate conclusions may be drawn from the research.

This research needs to be replicated over longer time periods and a diverse set of geographic areas to determine if these results are robust over differing situations. In addition, further research should incorporate other variables, such as soil type, soil fertility, chemical applications and weather conditions. This will determine if temporal grid yield variability is related to these additional pieces of site-specific information.

Endnotes

1. The Pearson correlations were also calculated for the actual grid yields. These were found to be nearly the same as the Pearson correlations for the standardized yield grid data.
2. To check whether the Pearson correlations are sensitive to the number of surrounding grids, Pearson correlations were calculated only for those grids that had four surrounding grids. While the magnitude and significance of some of the correlations changed, the story that emerges from the correlations does not change.

References

1. Eghball, Bahman and Gary E Varvel. "Fractal Analysis of Temporal Yield Variability of Crop Sequences: Implications for Site-Specific Management." *Agronomy Journal.* 89(Nov/Dec 1997): 851-855.

2. Jaynes, Dan B. and Thomas S. Colvin. "Spatiotemporal Variability of Corn and Soybean Yield." *Agronomy Journal.* 89(1997): 30-37.

3. Krill, Thomas L. Site Specific Management: What it is, What We've Done, and Challenges for the Future. Seminar. The Ohio State University Extension. Columbus, Ohio. 1997.

4. Lowenberg-DeBoer, Jess. "Risk Management Potential of Precision Farming Technologies." *Journal of Agricultural and Applied Economics.* To be published in a forthcoming journal.

5. Weiss, Michael D. "Precision Farming and Spatial Economic Analysis: Research Challenges and Opportunities." *American Journal of Agricultural Economics.* 78(December 1996): 1275-1280.

Table 1: Selected Summary Data for Fields 1, 2 and 3 by Year, Henry County, Ohio, 1995-1998

		Statistical Measures				
Crop	Year	Mean Grid Yield	Median Grid Yield	Standard Deviation of Grid Yields	Maximum Grid Yield	Minimum Grid Yield
			Field 1			
Beans	1998	48.3	47.5	10.4	66.0	32.0
Corn	1997	120.1	119.0	13.1	160.0	90.0
Corn	1996	135.2	136.0	12.0	166.0	104.0
Beans	1995	39.1	38.0	12.3	67.0	38.0
			Field 2			
Corn	1998	139.8	139.0	17.3	172.0	139.0
Beans	1997	31.6	30.5	9.3	63.0	30.5
Corn	1996	124.1	124.0	8.9	146.0	124.0
Beans	1995	39.1	38.0	12.3	67.0	38.0
			Field 3			
Beans	1998	51.0	49.0	10.2	112.0	40.0
Corn	1997	130.4	131.0	14.5	167.0	89.0
Beans	1996	32.9	32.0	13.3	68.0	6.0
Corn	1995	124.0	125.0	16.8	169.0	70.0

SOURCE: Original Data and Analysis

Table 2: Pearson Correlation Matrix Between a Standardized Yield for a Grid for a Year, and the Standardized Yield for the Grid for a Previous Year, Henry County, Ohio, 1995-1998*

Field 1			
Standardized Yield for a Grid in Year:	Standardized Yield for a Grid in Year:		
	1997	1996	1995
1998	-0.055 (0.531)	0.006 (0.062)	0.004 (0.043)
1997		-0.072 (0.660)	0.003 (0.035)
1996			-0.126 (0.905)
Field 2			
Standardized Yield for a Grid in Year:	Standardized Yield for a Grid in Year:		
	1997	1996	1995
1998	0.165 (0.948)	0.031 (0.284)	-0.137 (0.893)
1997		0.005 (0.046)	-0.213 (0.988)
1996			0.120 (0.840)
Field 3			
Standardized Yield for a Grid in Year:	Standardized Yield for a Grid in Year:		
	1997	1996	1995
1998	0.101 (0.770)	0.056 (0.496)	0.013 (0.125)
1997		-0.224 (0.993)	0.013 (0.126)
1996			0.082 (0.670)

*The statitiscal confidence level is presented in parentheses.
SOURCE: Original Data and Analysis

Table 3: Pearson Correlation Matrix Between Standardized Yield for a Grid for a Year, and the Standardized Yield for the Grid Averaged Across Previous Years, Henry County, Ohio, 1995-1998*

Field 1		
Standardized Yield for a Grid in Year:	Average Standardized Yield for the Grid in Years: 1995-1997	1995-1996
1998	-0.028 (0.286)	0.008 (0.080)
1997		-0.052 (0.509)
Field 2		
Standardized Yield for a Grid in Year:	Average Standardized Yield for the Grid in Years: 1995-1997	1995-1996
1998	0.034 (0.308)	-0.071 (0.594)
1997		-0.139 (0.898)
Field 3		
Standardized Yield for a Grid in Year:	Average Standardized Yield for the Grid in Years: 1995-1997	1995-1996
1998	0.103 (0.779)	0.047 (0.425)
1997		-0.143 (0.913)

*The statitiscal confidence level is presented in parentheses.
SOURCE: Original Data and Analysis

Table 4: Pearson Correlation Matrix Between a Standardized Yield for a Grid for a Year, and the Average Standardized Yield for the Grid and Surrounding Grids for a Previous Year, Henry County, Ohio, 1995-1998*

	Field 1		
Standardized Yield for a Grid in Year:	Average Standardized Yield for a Grid & Surrounding Grids in Year:		
	1997	1996	1995
1998	0.003 (0.027)	0.050 (0.421)	-0.040 (0.343)
1997		-0.067 (0.545)	-0.160 (0.927)
1996			0.024 (0.212)
	Field 2		
Standardized Yield for a Grid in Year:	Average Standardized Yield for a Grid & Surrounding Grids in Year:		
	1997	1996	1995
1998	0.138 (0.823)	-0.059 (0.437)	-0.243 (0.983)
1997		-0.181 (0.924)	-0.275 (0.994)
1996			-0.011 (0.087)
	Field 3		
Standardized Yield for a Grid in Year:	Average Standardized Yield for a Grid & Surrounding Grids in Year:		
	1997	1996	1995
1998	0.377 (0.999)	-0.169 (0.904)	0.058 (0.430)
1997		-0.233 (0.979)	-0.045 (0.341)
1996			0.073 (0.527)

*The statitiscal confidence level is presented in parentheses.

SOURCE: Original Data and Analysis

Table 5: Pearson Correlation Matrix Between Standardized Yield for a Grid for a Year, and the Standardized Yield for the Grid and Surrounding Grids Averaged Across Previous Years, Henry County, Ohio, 1995-1998*

Field 1		
Standardized Yield for a Grid in Year:	Average Standardized Yield for a Grid & Surrounding Grids in Years: 1995-1997	1995-1996
1998	0.010 (0.090)	0.009 (0.078)
1997		-0.159 (0.925)
Field 2		
Standardized Yield for a Grid in Year:	Average Standardized Yield for a Grid & Surrounding Grids in Years: 1995-1997	1995-1996
1998	-0.097 (0.657)	-0.202 (0.952)
1997		-0.298 (0.997)
Field 3		
Standardized Yield for a Grid in Year:	Average Standardized Yield for a Grid & Surrounding Grids in Years: 1995-1997	1995-1996
1998	0.136 (0.817)	-0.083 (0.585)
1997		-0.191 (0.941)

*The statitiscal confidence level is presented in parentheses.

SOURCE: Original Data and Analysis